**Source: Fraunhofer IIS**

**Title: JBM experiments for characterization testing**

## Document for: Discussion and Agreement

## Agenda Item: 7.5

1. Introduction

The IVAS-8b P-doc [1], "Test Plan for Characterization," outlines the allocation of three experiments to evaluate JBM, frame error (FE), and tandem coding, as detailed in Table 7. However, the specific conditions for the subjective P.800 experiments P.800-21, P.800-22, and P.800-23 are currently lacking in Annex F. This contribution aims to provide the necessary details for these experiments, with the main objective to test JBM functionality.

1. Discussion

In [2], a single delay and error profile, I1, comprising various impairments from existing profiles, was introduced and accepted for testing the characterization of JBM functionality. With the challenge of not having 5G profiles resolved, experiments to assess the functionality of the JBM implementation in both floating-point and fixed-point code can be designed. To avoid testing only one profile, variants with different offsets can be created. Therefore, as shown in the proposed tables below, two profile variants are used: I1.O1 and I1.O2.

In Table 7 of [1] it is foreseen, experiments P.800-21, P.800-22, and P.800-23 will evaluate both FE and tandem coding alongside JBM. Given the limited testing capacity and the complexity of testing tandem coding, which would also involve format conversion to mono, the source proposes not to include this aspect in the characterization tests. Thus, the primary objective of the aforementioned experiments should focus on testing IVAS JBM functionality, compared to IVAS decoding with FE, with error patterns derived from the profile variants I1.O1 and I1.O2, and clean channel conditions.

1. Proposal

As outlined in Table 7 of [1], the input formats designated for testing JBM functionality include stereo, object-based audio with 1-2 objects, and FOA. To ensure the robustness of the JBM implementation, both DTX and varying input loudness levels should be incorporated into the testing. These considerations are reflected in the detailed conditions of the experiments described below.

* 1. Experiment P800-21: JBM with Stereo

Tables F.21.1 to F.21.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Music and Mixed content categories, respectively.

Table F21.1: Conditions for Experiment P800-21

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96 kbps |
| DTX | DTX on |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15 dB for cat 3,4 |
| Error Conditions | JBM Profiles I1. O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRUESDRU | Q = 12, 17, 22, 27 dB $ $*α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation:  | Cat. 1-4: Model-based relying on convolution of raw mono clean speech sentences with Room Impulse Responses respective to various talker positions relative to a capture point as described in the ITU-T Reverberation Tool [18] and impulse responses provided by MC.Cat. 5-6: Pre-produced content |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to SWB for categories 1-4, up to FB for categories 5-6 |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content |
| Number of categories | 6 Different environments and talker interactions  |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |
|  |  |

Table F.21.2 : Preliminaries for Experiment P800-21

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c31 |  | IVAS FL  | 24.4 | I1.O1 | on |
| 2 | c24 |  | IVAS FL  | 48 | no error | on |
| 3 | c06 |  | ESDRU $α$ = 0.7 | - | - | - |
| 4 | c30 |  | IVAS FL  | 96 | Error I1.O2 | on |
| 5 | c09 |  | ESDRU $α$ = 0.1 | - | - | - |
| 6 | c35 |  | IVAS FL  | 48 | I1.O2 | on |
| 7 | c03 |  | MNRU Q = 27 dB | - | - | - |
| 8 | c01 |  | Reference | - | - | - |
| 9 | c26 |  | IVAS FL | 24.4 | Error I1.O1 | on |
| 10 | c07 |  | ESDRU $α$ = 0.5 | - | - | - |
| 11 | c05 |  | MNRU Q = 12 dB | - | - | - |
| 12 | c25 |  | IVAS FL  | 96 | no error | on |

Table F.21.3: Test conditions for Experiment P800-21,
stereo speech and music and mixed content under impaired and clean channel conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **DTX** | **Profile** |
| c01 | Reference | - | - | - |
| c02 | MNRU Q = 27 dB | - | - | - |
| c03 | MNRU Q = 22 dB | - | - | - |
| c04 | MNRU Q = 17 dB | - | - | - |
| c05 | MNRU Q = 12 dB | - | - | - |
| c06 | ESDRU $α=0.7$ | - | - | - |
| c07 | ESDRU $α=0.5$ | - | - | - |
| c08 | ESDRU$α=0.3$ | - | - | - |
| c09 | ESDRU$α=0.1$ | - | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 | on | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | on | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | on | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | on | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | on | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | on | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | on | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | on | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | on | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | on | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | on | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | on | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | on | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | on | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | on | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | on | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | on | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | on | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | on | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | on | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | on | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | on | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | on | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | on | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | on | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | on | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | on | I1. O2 |

Table F.2.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Room***  | ***Reverb***  | ***Microphone Setup*** | ***Background*** | ***SNR******[dB]*** | ***Overtalk [s](1*** | ***Bandwidth***  | ***Talker positions(2*** | ***Talker selection by panel*** |
| cat 1 | small | anechoic | M-S | Low level idle noise | 45 | 1 | Max available up to SWB | 1-75-32-64-13-47-2 | P1: f1m1P2: m2f2P3: f3m3P4: m1f1P5: f2m2P6: m3f3 |
| cat 2 | large | echoic | A-B (150 cm) | Low level idle noise | 45 | -1 | max available up to SWB | 5-111-63-75-89-710-9 | P1: m3f3P2: f1m1P3: m2f2P4: f3m3P5: m1f1P6: f2m2 |
| cat 3 | small | echoic | Binaural | office | 15 | 1 | max available up to SWB | 1-75-32-64-13-47-2 | P1: f2m2P2: m3f3P3: f1m1P4: m2f2P5: f3m3P6: m1f1 |
| cat 4 | car | car | A-B Cardioid pair 20 cm | car | 15 | -1 | Max available up to SWB | Driver-PassengerBackRight-DriverDriver-BackCenterBackLeft-DriverBackRight-BackLeftBackCenter-BackRight | P1: m1f1P2: f2m2P3: m3f3P4: f1m1P5: m2f2P6: f3m3 |

Table F.2.5: Mixed content and music categories

|  |  |
| --- | --- |
| **Category**  | **Type** |
| cat 5 | mixed content |
| cat 6 | music |

**Notes:**

**(1** Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

**(2** The talker positions are part of the scene definition of the different categories. They correspond to the talker positions as depicted in Figures 14.3 and 14.5 of [18] for the large and the small room, respectively.

* 1. Experiment P800-22: JBM with 1-2 objects

Tables F.22.1 to F.22.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Speech with background and Music and mixed content categories, respectively.

Table F.22.1: Conditions for Experiment P800-22

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96kbps |
| DTX | DTX off |
| Input level | -16, -26, -36 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise  |
| Error Conditions | JBM Profiles I1.O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRUESDRU | Q = 18, 24, 30, 36 dB $ $*α* = 0.1, 0.3, 0.5, 0.7 |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation:  | Cat. 1-2: Defined scenes, 1 ISMCat. 3-4: Defined scenes, 2 ISMsCat. 5, 6: Pre-produced content |
| Binaural renderer | ISM to binaural internal rendering  |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), music and mixed content, speech and background |
| Number of categories | 6  |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 4.2.1.1  |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.22.2: Preliminaries for Experiment P800-22

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c31 |  | IVAS FL  | 24.4 | I1.O1 |
| 2 | c24 |  | IVAS FL  | 48 | no error |
| 3 | c06 |  | ESDRU $α$ = 0.7 | - | - |
| 4 | c30 |  | IVAS FL  | 96 | Error I1.O2 |
| 5 | c09 |  | ESDRU $α$ = 0.1 | - | - |
| 6 | c35 |  | IVAS FL  | 48 | I1.O2 |
| 7 | c03 |  | MNRU Q = 27 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c26 |  | IVAS FL | 24.4 | Error I1.O1 |
| 10 | c07 |  | ESDRU $α$ = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 12 dB | - | - |
| 12 | c25 |  | IVAS FL  | 96 | no error |

Table F.22.3: Test conditions for Experiment P800-22,
speech under clean and impaired channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **Profile** |
| c01 | Reference | - | - |
| c02 | MNRU Q = 36 dB | - | - |
| c03 | MNRU Q = 30 dB | - | - |
| c04 | MNRU Q = 24 dB | - | - |
| c05 | MNRU Q = 18 dB | - | - |
| c06 | ESDRU $α=0.7$ | - | - |
| c07 | ESDRU $α=0.5$ | - | - |
| c08 | ESDRU$α=0.3$ | - | - |
| C09 | ESDRU$α=0.1$ | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | I1. O2 |

**Scene definitions categories 1-2**

A leading and trailing silence is present for each sample, in accordance with IVAS-7b. The metadata corresponds to the whole duration of the samples. This means that for moving objects, only a part of the trajectory corresponds to active speech. The following scenes are used:

1. Talker sitting at a table (elevation 0°), at different azimuths.
2. Standing talker (elevation 35°), at different azimuths.
3. Smaller talker (child) walking around a table in the positive sense (counterclockwise), elevation 0°. Azimuth varies continuously for the sentence pair.
4. Adult talker walking around a table in the negative sense (clockwise), elevation 35°. Azimuth varies continuously for the sentence pair.
5. Elevation displacement: Elevation varies continuously for the sentence pair. Azimuth is constant for a sentence pair, but different for each sentence pair.
6. Azimuth and elevation displacement: Azimuth and elevation vary continuously.

Each of the sentences uttered by a certain talker is encoded using different scene. Allocation of scenes to each panel is given in the Table F.9.4.

**Scene definitions categories 3-4**

The listening database consists of artificially created spatial audio samples from monophonic clean speech recordings where always 1 female and 1 male talker are combined in conversation-like scenarios following the Scene descriptions below.

A leading and trailing silence is present for each artificially created spatial audio sample, in accordance with IVAS-7b. The metadata corresponds to the whole duration of the sample. This means that for moving objects, only a part of the trajectory corresponds to active speech.

In one half of the samples, the 2nd talker’s utterance follows the 1st talker’s utterance simulating natural conversation. The gap between the utterances is set to 1 s. In the other half of the samples, the situation is similar, but the utterances partially overlap. The targeted overlap is also 1 s. Non-overlapping sentence pairs are used for Scenes a), c), and e) as described below. Overlapping sentence pairs are used for Scenes b), d), and f). The following scenes are used:

1. Two talkers sitting at a table (elevation 0°), at different azimuths. To increase positional variation, both the absolute azimuths and the difference of the azimuths of both talkers vary for each sentence pair. Non-overlapping utterances.
2. Two standing talkers (elevation 35°), at different azimuths. To increase positional variation, both the absolute azimuths and the difference of the azimuths of both talkers vary for each sentence pair. Overlapping utterances.
3. One talker sitting at a table (elevation 0°), second talker standing beside the table (elevation 45°). Non-overlapping utterances.
4. One talker sitting at a table (elevation 0°), second talker walking around the table (elevation 45°). The azimuth of the 2nd talker varies continually. Overlapping utterances.
5. Two talkers walking side-by-side around the table (elevation 45°). The azimuth is the same for both talkers and varies continually. Non-overlapping utterances.
6. Two talkers walking around the table in opposite directions (elevation 30°), starting at the same position. Azimuths of both talkers vary continually. Overlapping utterances.

The following table lists the test Categories corresponding to different talkers or talker pairs. Each of the sentence pairs uttered by a certain talker or talker pair is associated to a different scene.

Table F.22.4: Allocation of scenes for each talker or talker pair (category cat 1, cat 2, cat 3, cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Scene*** | ***Talker initial elevation*** | ***Elevation change*** | ***Talker initial azimuth*** | ***Azimuth change(2*** | ***Panel*** |
| ***cat 1:****M1*  | abefcd | 0°35°-90°35°0°35° | staticstatic0.3°/ frame-0.2°/ frame staticstatic | 270°180°120°0°240°180° | staticstaticstatic0.5°/ frame1°/ frame-1°/ frame | P1P2P3P4P5P6 |
| ***cat 2:****F1* | fcdabe | 35°0°35°0°35°-90° | -0.2°/ framestaticstaticstaticstatic0.3°/ frame | 300°60°120°60°300°60° | 0.5°/ frame1°/ frame-1°/ framestaticstaticstatic | P1P2P3P4P5P6 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Scene*** | ***Overtalk******[s]**(1*** | ***1st talker elevation*** | ***2nd talker elevation*** | ***1st talker initial azimuth*** | ***1st talker azimuth change(2*** | ***2nd talker initial azimuth*** | ***2nd talker azimuth change(2*** | ***Panel*** |
| ***cat 3:****M2 + F2* | *a**b**c**d**e**f* | *-1**1**-1**1**-1**1* | *0°**35°**0°**0°**45°**30°* | *0°**35°**45°**45°**45°**30°* | *0°**10°**20°**200°**340°**120°* | *static**static**static**static**-1°/ frame**1°/ frame* | *50°**110°**170°**30°**340°**120°* | *static**static**static**-1°/ frame**-1°/ frame**-1°/ frame* | *P1**P2**P3**P4**P5**P6* |
| ***cat 4:****M3 + F3* | *d**e**f**a**b**c* | *1**-1**1**-1**1**-1* | *0°**45°**30°**0°**35°**0°* | *45°**45°**30°**0°**35°**45°* | *50°**130°**300°**30°**40°**50°* | *static**1°/ frame**1°/ frame**static**static**static* | *180°**130°**300°**230°**290°**350°* | *1°/ frame**1°/ frame**-1°/ frame**static**static**static* | *P1**P2**P3**P4**P5**P6* |

**Notes:**

(1Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

(2 The positive sense for azimuth is counterclockwise

Table F.9.5: Music and mixed content and Speech and background categories

|  |  |
| --- | --- |
| **Category**  | **Type** |
| cat 5 | Music and mixed content (1 object) |
| cat 6 | speech + background (2 objects) |

* 1. Experiment P800-23: JBM with FOA

Tables F.23.1 to F.23.5 show conditions to be used for this experiment, list of preliminaries, full list of conditions, and definition of Speech categories, and Mixed content and Generic audio categories, respectively.

Table F.23.1: Conditions for Experiment P800-23

|  |  |
| --- | --- |
| **Main Codec Conditions** |  |
| Candidate | CuT IVAS FX, CuT IVAS FL |
| Bitrates | 24.4, 48, 96 kbps |
| DTX | DTX off |
| Input level | -26 LKFS |
| Input frequency mask | 20KBP |
| Noise | No noise for cat 1,2,5,6, 15 dB for cat 3,4 |
| Error Conditions | JBM Profiles I1.O1, I1.O2 (offset O1 is a random number and O2=(O1+4000)%8000)Error patterns Error I1.O1, Error I1.O2 (error pattern files derived from the respective delay and error profiles) |
| **References** |  |
| Direct | -26 LKFS |
| P.50 MNRUESDRU | Q = xx, xx, xx, xx dB $ $*α* = xx, xx, xx |
| Input frequency mask | 20KBP |
| **Common Conditions** |  |
| Test item generation: pre-processing incl. spatialization | Cat. 1-4: Model-based relying on convolution of raw mono clean speech sentences convolved with (FOA) Spatial Room Impulse Responses respective to various talker positions relative to a capture point and spatial (FOA) ambient noise mixing.Cat. 5-6: Pre-produced content |
| Binaural renderer | FOA to binaural internal rendering  |
| Audio sampling Frequency/bandwidth | 48 kHz/maximum available audio bandwidth up to FB |
| Kind of samples | Sentence pairs uttered by different talkers and genders (3 male and 3 female), mixed content, generic audio |
| Number of categories | 6 Different environments (with or without background) and talker interactions  |
| Number of samples | 6 + 1 (preliminaries) samples per category |
| Listening Level | 73 dB SPL |
| Listeners | Naïve listeners |
| Randomizations | 6 panels of 5 listeners |
| Rating Scale | Following clause 4.2.1.1 |
| Listening System | Headphones, in accordance with clause 4.4 |
| Listening Environment | No room noise |

Table F.23.2: Preliminaries for Experiment P800-23

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c31 |  | IVAS FL | 24.4 | I1.O1 |
| 2 | c24 |  | IVAS FL  | 48 | no error |
| 3 | c06 |  | ESDRU $α$ = 0.7 | - | - |
| 4 | c30 |  | IVAS FL  | 96 | Error I1.O2 |
| 5 | c09 |  | ESDRU $α$ = 0.1 | - | - |
| 6 | c35 |  | IVAS FL  | 48 | I1.O2 |
| 7 | c03 |  | MNRU Q = 27 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c26 |  | IVAS FL  | 24.4 | Error I1.O1 |
| 10 | c07 |  | ESDRU $α$ = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 12 dB | - | - |
| 12 | c25 |  | IVAS FL  | 96 | no error |

Table F.23.3: Test conditions for Experiment P800-23,
speech and mixed-music under impaired and clean channel conditions

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Condition** | **Bitrate [kbps]** | **Profile** |
| c01 | Reference | - | - |
| c02 | MNRU Q = xx dB | - | - |
| c03 | MNRU Q = xx dB | - | - |
| c04 | MNRU Q = xx dB | - | - |
| c05 | MNRU Q = xx dB | - | - |
| c06 | ESDRU $α=xx$ | - | - |
| c07 | ESDRU $α=xx$ | - | - |
| c08 | ESDRU$α=xx$ | - | - |
| c09 | ESDRU$α=xx$ | - | - |
| c10 | IVAS FL enc / FX dec | 24.4 | No error |
| c11 | IVAS FL enc / FX dec | 48.0 | No error |
| c12 | IVAS FL enc / FX dec | 96.0 | No error |
| c13 | IVAS FL enc / FX dec | 24.4 | Error I1. O1 |
| c14 | IVAS FL enc / FX dec | 48.0 | Error I1. O1 |
| c15 | IVAS FL enc / FX dec | 24.4 | Error I1. O2 |
| c16 | IVAS FL enc / FX dec | 48.0 | Error I1. O2 |
| c17 | IVAS FL enc / FX dec | 24.4 | I1. O1 |
| c18 | IVAS FL enc / FX dec | 48.0 | I1. O1 |
| c19 | IVAS FL enc / FX dec | 96.0 | I1. O1 |
| c20 | IVAS FL enc / FX dec | 24.4 | I1. O2 |
| c21 | IVAS FL enc / FX dec | 48.0 | I1. O2 |
| c22 | IVAS FL enc / FX dec | 96.0 | I1. O2 |
| c23 | IVAS FL enc / FL dec | 24.4 | No error |
| c24 | IVAS FL enc / FL dec | 48.0 | No error |
| c25 | IVAS FL enc / FL dec | 96.0 | No error |
| c26 | IVAS FL enc / FL dec | 24.4 | Error I1. O1 |
| c27 | IVAS FL enc / FL dec | 48.0 | Error I1. O1 |
| c28 | IVAS FL enc / FL dec | 24.4 | Error I1. O2 |
| c29 | IVAS FL enc / FL dec | 48.0 | Error I1. O2 |
| c30 | IVAS FL enc / FL dec | 96.0 | Error I1. O2 |
| c31 | IVAS FL enc / FL dec | 24.4 | I1. O1 |
| c32 | IVAS FL enc / FL dec | 48.0 | I1. O1 |
| c33 | IVAS FL enc / FL dec | 96.0 | I1. O1 |
| c34 | IVAS FL enc / FL dec | 24.4 | I1. O2 |
| c35 | IVAS FL enc / FL dec | 48.0 | I1. O2 |
| c36 | IVAS FL enc / FL dec | 96.0 | I1. O2 |

Table F23.4: Clean and noisy speech categories and scene definitions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Environment(1*** | ***Background(2***  | ***SNR [dB]*** | ***Overtalk [s](3*** | ***Bandwidth***  | ***Talker positions(4*** | ***Talker selection by panel*** |
| *cat 1* | *room\_1\_FOA*  | *room\_1\_cleanbg\_FOA* | *45* | *1* | *Max*  |  | *P1: f1m1P2: m2f2P3: f3m3P4: m1f1P5: f2m2P6: m3f3* |
| *cat 2* | *room\_4\_FOA*  | *Room\_[1/4]\_cleanbg\_FOA* | *45* | *-1* | *Max*  |  | *P1: m3f3P2: f1m1P3: m2f2P4: f3m3P5: m1f1P6: f2m2* |
| *cat 3* | *out\_[X]\_FOA* | *[park\_1\_bg\_FOA / nature\_1\_bg\_FOA / event\_1\_bg\_FOA / street\_[1/2]\_bg\_FOA]* | *15* | *1* | *Max* |  | *P1: f2m2P2: m3f3P3: f1m1P4: m2f2P5: f3m3P6: m1f1* |
| *cat 4* | *room\_[X]\_FOA* | *[cafeteria\_1\_bg\_FOA / mall\_1\_bg\_FOA/ office[1/2]\_bg\_FOA]* | *15* | *-1* | *Max* |  | *P1: m1f1P2: f2m2P3: m3f3P4: f1m1P5: m2f2P6: f3m3* |

Table F23.5: Mixed content and Generic audio categories

|  |  |
| --- | --- |
| **Category**  | **Type** |
| cat 5 | mixed content |
| cat 6 | generic audio |

**Notes:**

**(1** Editor’s note: The specific room/environment characteristic and resulting reverb characteristic will be defined by the choice of the specific Spatial Room Impulse Responses used in the convolution process with the raw mono sentences, according to the pertaining stipulations of the test plan IVAS-8b.

**(2** Editor’s note: Background is defined by the chosen background noise file according to the pertaining stipulations of the test plan IVAS-8b. Backround name ‘clean\_bg\_[X]\_FOA’ indicates a low-noise background corresponding to environment [X], e.g., with low air-conditioning/fan noise.

**(3** Overtalk [s] means the duration in seconds by which the two sentences in the sound item uttered by different talkers are overlapping. A negative number means that there is a corresponding pause between the two sentences.

**(4** Editor’s note: The talker positions are part of the scene definition of the different categories. They should be chosen in a way from the available set of SRIRs for the used room making sure that there is a good coverage of different possible positions. Different selections should be made for the different listener panels.

1. Conclusion

The source kindly requests agreement on the proposed experiment descriptions for inclusion in Annex F of [1].

1. References
2. S4-250276 - IVAS Permanent Document IVAS-8b: Test Plan for Characterization Phase, v0.5.0
3. S4-250217 - [IVAS\_Codec\_Ph2] Delay and Error Profile for Characterization Testing, Fraunhofer IIS